

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number
WO 01/23610 A2(51) International Patent Classification⁷: C12Q 1/68 (74) Agent: GILL JINNINGS & EVERY; Broadgate House, 7 Eldon Street, London EC2M 7LH (GB).

(21) International Application Number: PCT/GB00/03734

(22) International Filing Date:
29 September 2000 (29.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9923084.9 29 September 1999 (29.09.1999) GB
0002216.0 31 January 2000 (31.01.2000) GB

(71) Applicant (for all designated States except US): SOLEXA LTD: [GB/GB]; 38 Jermyn Street, London SW1Y 6DN (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): BALASUBRAMANIAN, Shankar [GB/GB]; University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/23610 A2

(54) Title: POLYNUCLEOTIDE SEQUENCING

(57) Abstract: A method for sequencing polynucleotides involves sequential addition of detectably labelled bases. A labelled base is incorporated by a polymerase onto a nascent strand, and, after detection, the polymerase is induced to replace the terminal labelled base with a corresponding unlabelled base, thereby permitting further sequencing to occur. The exonuclease activity of many polymerase enzymes is used, to remove the terminal base.

POLYNUCLEOTIDE SEQUENCING

Field of the Invention

This invention relates to the sequencing of polynucleotides. In particular, this invention discloses methods for determining the sequence of polynucleotides arrayed 5 on a solid support.

Background to the Invention

Advances in the study of molecules have been led, in part, by improvement in technologies used to characterise the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA has benefited from developing 10 technologies used for sequence analysis and the study of hybridisation events.

An example of the technologies that have improved the study of nucleic acids, is the development of fabricated arrays of immobilised nucleic acids. These arrays consist typically of a high-density matrix of polynucleotides immobilised onto a solid support material. Fodor *et al*, Trends in Biotechnology (1994) 12:19-26, describes 15 ways of assembling the nucleic acids using a chemically sensitized glass surface protected by a mask, but exposed at defined areas to allow attachment of suitably modified nucleotide phosphoramidites. Fabricated arrays may also be manufactured by the technique of "spotting" known polynucleotides onto a solid support at predetermined positions (e.g. Stimpson *et al*, PNAS (1995) 92:6379-6383).

20 A further development in array technology is the attachment of the polynucleotides to the solid support material to form single molecule arrays. Arrays of this type are disclosed in WO-A-00/06770. The advantage of these arrays is that reactions can be monitored at the single molecule level and information on large numbers of single molecules can be collated from a single reaction.

25 For DNA arrays to be useful, the sequences of the molecular must be determined. US-A-5302509 discloses a method to sequence polynucleotides immobilised on a solid support. The method relies on the incorporation of 3'-blocked bases A, G, C and T having a different fluorescent label to the immobilised polynucleotide, in the presence of DNA polymerase. The polymerase incorporates a 30 base complementary to the target polynucleotide, but is prevented from further addition by the 3'-blocking group. The label of the incorporated base can then be determined and the blocking group removed by chemical cleavage to allow further polymerisation to occur. However, the need to remove the blocking groups in this manner is time-consuming and must be performed with high efficiency.

Similarly, EP-A-0640146 discloses a polymerisation-based technique for sequencing DNA. The technique again requires removal of a blocking group prior to subsequent incorporation of nucleotides.

Summary of the Invention

5 In the general method of the invention, a target polynucleotide sequence can be determined by generating its complement using the polymerase reaction to extend a suitable primer, and characterising the successive incorporation of bases that generate the complement. The target sequence is, typically, immobilised on a solid support. Each of the different bases A, T, G or C is then brought, by sequential 10 addition, into contact with the target, and any incorporation events detected via a suitable label attached to the base. In contrast to the prior art methods, the present invention requires the presence of a polymerase enzyme that retains a 3' to 5' exonuclease function, which is induced to remove an incorporated labelled base after 15 detection of incorporation. A corresponding non-labelled base may then be incorporated into the complementary strand to allow further sequence determinations to be made. Repeating the procedure allows the sequence of the complement to be identified, and thereby the target sequence also.

The use of the polymerase enzyme's exonuclease function in this way is a characteristic feature of the invention. It permits repeated incorporation of labelled 20 bases to take place, without the requirement for separate steps of chemical cleavage or photobleaching.

Accordingly, a method for determining the sequence of a target polynucleotide on an array, comprises the steps of:

- (i) contacting the array with one or more detectably-labelled bases A, 25 T, G and C, under conditions that permit the polymerisation reaction to occur, to thereby incorporate a labelled base into a strand complementary to the target;
- (ii) removing non-incorporated bases and detecting an incorporation event;
- 30 (iii) optionally repeating steps (i) and (ii) with one or more additional labelled bases, to determine a partial sequence;
- (iv) contacting the array of step (iii) with a DNA polymerase having 3' to 5' exonuclease activity, under conditions whereby the polymerase

cleaves the labelled base(s) and incorporates corresponding non-labelled base(s); and

(v) repeating steps (i) - (iv) sequentially, to determine the sequence.

Sequencing the polynucleotides on the array makes it possible to form a 5 spatially addressable array. This may then be used for many different applications, including genotyping studies and other characterisation experiments.

The method of the present invention may be automated to produce a very efficient and fast sequence determination.

Description of the Drawings

10 Figure 1 illustrates the gradual replacement of a terminal labelled base with an unlabelled base, using a polymerase enzyme.

Description of the Invention

15 The method according to the invention, for determining the sequence of the arrayed polynucleotides, is carried out by contacting the array separately with the different bases to form the complement to that of the target polynucleotide, and 20 detecting incorporation. The method makes use of polymerisation, whereby a polymerase enzyme extends the complementary strand by incorporating the correct base complementary to that on the target. The polymerisation reaction also requires a specific primer to initiate polymerisation.

25 For each cycle, the incorporation of a labelled base is carried out by the polymerase enzyme, and the incorporation event determined. Many different polymerase enzymes are now known to comprise an exonuclease function which is used to remove mismatches or mutations during normal DNA replication. This function is exploited in the present invention to remove the incorporated labelled base (or bases) after detection, permitting further sequence determinations to be made.

30 In the context of the invention, reference to the bases A, T, G and C is taken to be a reference to the deoxynucleoside triphosphates adenosine, thymidine, guanosine and cytidine, and to functional analogs thereof, including the chain termination dideoxynucleoside triphosphates which may be used as the labelled bases for the initial incorporation event.

The terms "arrayed polynucleotides" and "polynucleotide arrays" are used herein to define an array of polynucleotides that are immobilised on a solid support material. The polynucleotides may be immobilised to the solid support through a linker

molecule, or may be attached to a particle, e.g. a microsphere, which is itself attached to a solid support material.

The polynucleotides may be attached to the solid support by recognised means, including the use of biotin-avidin interactions. Methods for immobilising polynucleotides on a solid support are well known in the art, and include lithographic techniques and "spotting" individual polynucleotides in defined positions on a solid support. Suitable solid supports are known in the art, and include glass slides, ceramic and silicon surfaces and plastics materials. The support is usually a flat surface. In one embodiment, the polynucleotides are attached to the solid support via microscopic beads (microspheres), which may in turn be attached to the solid support by known means. The microspheres may be of any suitable size, typically in the range of from 10 nm to 100 nm in diameter. Attachment via microspheres allows discrete regions of polynucleotides to be easily generated on the array. Each microsphere may have multiple copies of a polynucleotide attached, and each microsphere can be resolved individually to determine incorporation events. Preferably, the arrays that are used are single molecule arrays that comprise polynucleotides in distinct optically resolvable areas, e.g. as disclosed in WO-A-00/06770.

The sequencing method may be carried out on both single molecule and multi molecule arrays, i.e. arrays of distinct individual molecules and arrays of distinct regions comprising multiple copies of one individual molecule. When multi-molecule arrays are used, it may be preferable to use a mixture of labelled and non-labelled base in the initial incorporation step. Diluting the concentration of labelled base in this way ensures that not every complementary strand incorporates a labelled base, and therefore distinct labels can be resolved within the relatively high densities of the multi molecule arrays. Single molecule arrays allow each individual polynucleotide to be resolved separately. The use of single molecule arrays is preferred. Sequencing single molecule arrays allows a spatially addressable array to be formed.

The term "spatially addressable" is used herein to describe how different molecules may be identified on the basis of their position on an array.

The method makes use of the polymerisation reaction to generate the complementary sequence of the target. The conditions necessary for polymerisation to occur will be apparent to the skilled person.

The polymerase required to carry out the incorporation of the labelled bases into the complementary strand (step (i)) does not need to have an exonuclease

function and indeed it is preferred if it does not have this function. The polymerase may be removed from the array by a washing step, and a suitable polymerase, comprising an exonuclease activity, brought into contact with the array in a subsequent step, for example on addition of the non-labelled bases. Suitable 5 polymerase enzymes will be known to those skilled in the art, and include DNA polymerase I, the Klenow fragment, DNA polymerase III and T4 or T7 DNA polymerase.

To carry out the polymerase reaction it will usually be necessary to first anneal 10 a primer sequence to the target polynucleotide, the primer sequence being recognised by the polymerase enzyme and acting as an initiation site for the subsequent extension of the complementary strand. The primer sequence may be added as a separate component with respect to the target polynucleotide. Alternatively, the 15 primer and the target polynucleotide may each be part of one single stranded molecule, with the primer portion forming an intramolecular duplex with a part of the target, i.e. a hairpin loop structure. This structure may be immobilised to the solid support at any point on the molecule.

Other conditions necessary for carrying out the polymerase reaction, including temperature, pH, buffer compositions etc., will be apparent to those skilled in the art.

This polymerisation step is allowed to proceed for a time sufficient to allow 20 incorporation of a base. Bases that are not incorporated are then removed, for example, by subjecting the array to a washing step, and detection of the incorporated labels may then be carried out.

In a preferred embodiment, the label is a fluorescent moiety and may be attached to the base in such a way to prevent further incorporation from occurring, i.e. 25 the base is a chain terminator. Many examples of fluorophores that may be used are known in the prior art e.g. tetramethylrhodamine (TMR). The attachment of a suitable fluorophore to a base can be carried out by conventional means. Suitably labelled bases are also available from commercial sources. When the label is a fluorophore, the fluorescence signal generated on incorporation may be measured by optical 30 means, e.g. by a confocal microscope.

Detection may be by conventional means, for example if the label is a fluorescent moiety, detection of an incorporated base may be carried out by using a confocal scanning microscope to scan the surface of the array with a laser, to image a fluorophore bound directly to the incorporated base. Alternatively, a sensitive 2-D

detector, such as a charge-coupled detector (CCD), can be used to visualise the individual signals generated. However, other techniques such as scanning near-field optical microscopy (SNOM) are available and may be used when imaging dense arrays. For example, using SNOM, individual polynucleotides may be distinguished 5 when separated by a distance of less than 100 nm, e.g. 10 nm to 10 μ m. For a description of scanning near-field optical microscopy, see Moyer *et al*, Laser Focus World (1993) 29:10. Suitable apparatus used for imaging polynucleotide arrays are known and the technical set-up will be apparent to the skilled person.

After the detection of an incorporation event, the exonuclease activity of the 10 polymerase enzyme is used to remove the labelled base. The polymerase is usually added separately after the washing step which may have removed the polymerase that incorporated the labelled base. The amount of polymerase required will depend on the amount of polynucleotides arrayed on the solid support, but can be determined readily by the skilled person. A suitable concentration will be 10 nM to 10 μ M. In 15 addition to the polymerase, an unlabelled base corresponding to that incorporated may be added. The unlabelled base will not be incorporated onto the complementary strand until removal of the labelled base. The unlabelled base may be a chain-terminator having a removable blocking group or ligand attached. The blocking group must be removed prior to repeating step (i). Suitable blocking groups are known in the 20 art and include photoactivatable ligands.

In the absence of further polymerisation, the polymerase will "switch on" its exonuclease activity to cleave the terminal base on the complementary strand. Cleavage will proceed until a suitable base is available for incorporation, when the 25 more favoured polymerisation reaction will occur. There is therefore an equilibrium reaction occurring with repeated cleavage and (if available) subsequent base incorporation at the terminal end of the complementary strand. Further cleavage, beyond the terminal base, is prevented by the preference of the polymerase for DNA synthesis.

The process of incorporating and removing labelled bases may then be 30 repeated using each of the different bases until the sequence has been determined.

The cleavage/polymerisation steps do not need to be repeated sequentially for each incorporated base. For example, it is possible that several bases are incorporated (and detected) prior to the cleavage step. To illustrate this, the following (template) sequence may be considered.

5'-ATG
3'-TACGTCTAT-5' (SEQ ID NO. 1)

A first labelled base C may be incorporated onto the template, and 5 incorporation detected. Further labelled bases A and G may then be added sequentially with detection. It will then be known that CAG forms a partial sequence of the growing strand. The labelled bases may then be removed by the exonuclease activity of a polymerase. The exonuclease activity may be induced by incubating the polymerase with the template in the presence of only the non-labelled base C. This 10 base cannot be incorporated onto the template and so the polymerisation reaction "stalls" and is replaced by the exonuclease activity. The labelled bases are removed until cleavage of the labelled C occurs, thereby permitting incorporation of the non-labelled C. The addition of the other non-labelled bases A and G can then be carried out to re-form the growing strand. The procedure may then be repeated for the next 15 part of the template sequence.

The following Example illustrates the invention, with reference to the accompanying drawing.

Example

A 20mer target polynucleotide (SEQ ID NO. 2) and a suitable (13mer) primer 20 molecule (SEQ ID NO. 3) were used having the sequences:

3'-TGGACGGCTGCGAATCGTC-5' (SEQ ID NO. 2)
5'-ACCTGCCGACGCT-3' (SEQ ID NO. 3)

Incorporation of the fluorescent base (TMR-dUTP; obtained from Nycomed 25 Amersham) was performed in a buffer of 50 mM Tris-HCl, pH 7.5, 10 mM NaCl, 2 mM DTT, 1 mM K₃PO₄, 0.1 mM EDTA, and 0.1 mg/ml BSA (100 µl total volume). The 13/20 mer duplex substrate and T4 DNA polymerase (lacking an exonuclease 30 function) were present at final concentrations of 100 nM and 150 nM, respectively. The polymerisation reaction was initiated by addition of a mixture of TMR-dUTP/MgCl₂ to final concentrations of 10 µM and 3 mM, respectively. After three minutes, the reaction was heated to 90°C (to inactivate the enzyme) and cooled slowly to reanneal the duplex.

After reannealing, BSA, dTTP and T4 DNA polymerase (having a 3' to 5' exonuclease function) were added to final concentrations of 0.1 mg/ml, 5 mM, and 500 nM, respectively. The reaction was allowed to proceed at 23°C, and then quenched, at various time points up to 30 minutes, using 0.5 M EDTA.

5 The reaction mixtures at each time point were resolved by electrophoresis and imaged using a Molecular Dynamics phosphoimager.

The results are shown in Fig. 1. In Fig. 1, lane 1 represents the 13/20mer standard at $t = 0$; lanes 2 and 3 represent TMR-dUTP incorporation with reaction quenching at $t = 1$ and 3 minutes, respectively; and lanes 4-9 represent quenching at 10 $t = 1, 3, 5, 10, 20$ and 30 minutes, respectively.

It can be seen that, during the time course of the reaction, the amount of 15 labelled 14mer (the 13mer and incorporated labelled base) is reduced as the labelled base is excised from the polynucleotide and an unlabelled base incorporated. This represents the "idling" of the polymerase and the "switching on" of the exonuclease function. This demonstrates that the exonuclease activity can be utilised to excise a terminal base and, provided that a suitable replacement base is present, the polynucleotide is maintained, ready for further sequencing to occur.

CLAIMS:

1. A method for determining the sequence of a target polynucleotide on an array, comprising the steps of:

- (i) contacting the array with one or more detectably-labelled bases A, T, G and C, under conditions that permit the polymerisation reaction to occur, to thereby incorporate a labelled base into a strand complementary to the target;
- 5 (ii) removing non-incorporated bases and detecting an incorporation event;
- 10 (iii) optionally repeating steps (i) and (ii) with one or more additional labelled bases, to determine a partial sequence;
- (iv) contacting the array of step (iii) with a DNA polymerase having 3' to 15 5' exonuclease activity, under conditions whereby the polymerase cleaves the labelled base(s) and incorporates corresponding non-labelled base(s); and
- (v) repeating steps (i) - (iv) sequentially, to determine the sequence.

2. A method according to claim 1, wherein the label is a fluorophore.

3. A method according to claim 1 or claim 2, wherein the labelled base is a chain terminator.

20 4. A method according to any preceding claim, wherein the non-labelled base comprises a removable blocking group which prevents further base incorporation from occurring, and wherein the blocking group is removed prior to repeating step (i).

5. A method according to any preceding claim, wherein the detection in step (ii) is carried out using optical means.

25 6. A method according to claim 5, wherein the optical means is a confocal microscope.

7. A method according to any preceding claim, wherein the polymerase of step (iv) is DNA polymerase I or III, the Klenow fragment, T4 or T7 polymerase.

8. A method according to any preceding claim, wherein step (i) is carried out 30 using a polymerase lacking exonuclease activity.

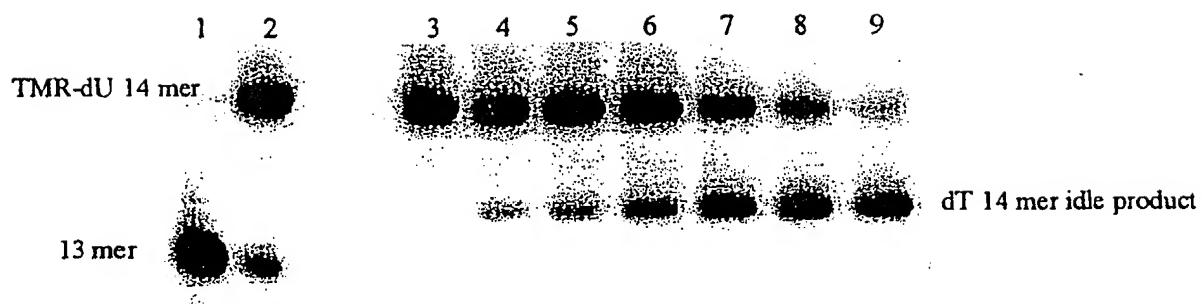


Fig. 1

SEQUENCE LISTING

<110> Solexa Ltd.

<120> Polynucleotide Sequencing

<130> REP06174WO

<140> (not yet known)

<141> 2000-09-28

<150> 9923084.9

<151> 1999-09-29

<150> 0002216.0

<151> 2000-01-31

<160> 3

<170> PatentIn Ver. 2.1

<210> 1

<211> 9

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Theoretical
polynucleotide

<400> 1

tacgtctat

9

<210> 2

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Template
polynucleotide

<400> 2

tggacggctg cgaatcgtc

19

<210> 3

<211> 13

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Primer
polynucleotide

<400> 3

acctgccgac gct

13